# Efficient algorithms for structured self-calibration problems

**Alistair B. Forbes**

*National Physical Laboratory, Teddington, Middlesex TW11 0LW, UK.*
alistair.forbes@npl.co.uk

**Abstract**

Self-calibration techniques have been used extensively in co-ordinate metrology. At their most developed, they are able to extract all systematic error behaviour associated with the measuring instrument as well as determining the geometry of the artefact being measured. However, this is generally at the expense of introducing extra parameters leading to moderately large observation matrices. Fortunately, these matrices tend to have sparse, block structure in which the nonzero elements are confined to much smaller submatrices. This structure can be exploited either in direct approaches in which QR factorisations are performed or in iterative algorithms which depend on matrix-vector multiplications. In this paper, we describe self-calibration approaches associated with high accuracy, dimensional assessment by co-ordinate measuring systems, highlighting how the associated optimisation problems can be presented compactly and solved efficiently. The self-calibration techniques lead to uncertainties significantly smaller than can be expected from standard methods.

## 1 Introduction

An important activity in metrology is the calibration of instruments and artefacts. Calibration defines a rule which converts the values output by the instrument's sensor(s) to values that can be related to the appropriate standard (SI or derived) units. Importantly, to these calibrated values it is required to assign uncertainties that reliably take into account the uncertainties of all quantities that have an influence. As a consequence, the size and complexity of the computational tasks associated with the data analysis can be significant, even for instruments that appear to be of simple design and operation. It is thus beneficial to design and implement algorithms that are efficient with respect to computation and memory. Fortunately, many of the calibration problems give rise to systems of equations with a well defined sparsity structure.

The rest of this paper is organised as follows. In Section 2 we review least squares approaches to calibration problems and go on to describe self-calibration problems in co-ordinate metrology in Section 3. Sections 4 and 5 describe solution methods for two types of sparsity structure. Our concluding remarks are given in Section 6.

## 2 Least squares solution to calibration problems

In many calibration problems, the observation equations involving measurements $y_i$

can be expressed as $y_i = \phi_i(\mathbf{a}) + \epsilon_i$, where $\phi_i$ is a function depending on parameters $\mathbf{a} = (a_1, \ldots, a_n)^{\mathrm{T}}$ specifying the behaviour of the instrument, and $\epsilon_i$ represents random measurement error. For a set of measurement data $\{y_i\}_1^m$, best estimates $\mathbf{a}^*$ of the calibration parameters $\mathbf{a}$ are determined by solving

$$\min_{\mathbf{a}} \sum_{i=1}^{m} f_i^2(\mathbf{a}) = \mathbf{f}^{\mathrm{T}}\mathbf{f}, \tag{2.1}$$

where $f_i(\mathbf{a}) = y_i - \phi_i(\mathbf{a})$. The most common approach to solving this problem is derived from the Gauss-Newton algorithm; see, for example, [5]. If $\mathbf{a}$ is an estimate of the solution and $J$ is the *Jacobian matrix* defined at $\mathbf{a}$ by $J_{ij} = \partial f_i/\partial a_j$, then an updated estimate of the solution is $\mathbf{a} + \mathbf{p}$, where $\mathbf{p}$ solves the Jacobian system

$$J\mathbf{p} = -\mathbf{f},$$

in the least squares sense. Starting with an appropriate initial estimate of $\mathbf{a}$, these steps are repeated until convergence criteria are met.

A numerically stable method of solving the Jacobian system is to find a factorisation $J = QR$, where $Q$ is an $m \times n$ orthogonal matrix and $R$ is an upper-triangular matrix of order $n$ (see, e.g., [1, 6]). The solution $\mathbf{p}$ is determined efficiently by solving the upper-triangular system

$$R\mathbf{p} = -Q^{\mathrm{T}}\mathbf{f},$$

using back substitution. The matrix $Q$ can be constructed using either Householder reflections, which process the Jacobian matrix a column at a time, or Givens plane rotations, which process the matrix row-wise. For either approach the orthogonal factorisation requires $O(mn^2)$ operations.

An alternative to the direct approaches to solve matrix equations is to use iterative procedures based on conjugate gradients. The advantage of these approaches is that they involve only matrix-vector multiplications and for sparse matrices these multiplications can be made efficient. In particular, the LSQR algorithm of Paigė and Saunders [7] implements an iterative approach to solving linear least squares problems.

Often, linear equality constraints on the parameters of the form $C^{\mathrm{T}}\mathbf{a} = \mathbf{c}$, where $C$ is an $n \times p$ matrix, $p < n$, are required to eliminate degrees of freedom in the problem. However, we can use orthogonal projections to eliminate these constraints. Suppose $C$ is of full column rank and has QR factorisation

$$C = \begin{bmatrix} V_1 & V_2 \end{bmatrix} \begin{bmatrix} S \\ 0 \end{bmatrix},$$

where $V_1$ and $V_2$, respectively, are the first $p$ and last $n - p$ columns of the orthogonal factor $V$. If $\mathbf{a}_0$ is a solution of $C^{\mathrm{T}}\mathbf{a} = \mathbf{c}$, then for any $(n-p)$-vector $\tilde{\mathbf{a}}$, $\mathbf{a} = \mathbf{a}_0 + V_2\tilde{\mathbf{a}}$ automatically satisfies the constraints and the optimisation problem can be reformulated as the unconstrained non-linear least squares problem

$$\min_{\tilde{\mathbf{a}}} \sum_{i=1}^{m} f_i^2(\mathbf{a}_0 + V_2\tilde{\mathbf{a}}),$$

involving the reduced set of parameters $\tilde{\mathbf{a}}$. We note that the associated Jacobian matrix is simply $\tilde{J} = JV_2$, where $J_{ij} = \partial f_i / \partial a_j$, as before.

Unfortunately, even if $J$ has structure $\tilde{J} = JV_2$ could be full. For indirect approaches, this is of little consequence since the matrix-vector multiplications can be formed in two stages (e.g., $\mathbf{y} = V_2\mathbf{x}$, $\mathbf{z} = J\mathbf{y}$) each of which can be implemented efficiently. For a direct approach, it may be possible to implement the constraints in such a way as to minimise the amount of fill-in during the orthogonal factorisation stage.

## 3 Self-calibration problems in co-ordinate metrology

Co-ordinate metrology is concerned with defining the geometry of two and three dimensional artefacts from measurements of the co-ordinates of points related to the surface of the artefacts. It is a key discipline in quality and process control in manufacturing industy. In a (conventional) co-ordinate measuring machine (CMM) with three mutually orthogonal linear axes, the position of the probe tip centre is inferred from scale readings on each of the three machine axes. In practice, CMMs have imperfect geometry with respect to the straightness of the axes, the squareness of pairs of axes and rotations describing roll, pitch and yaw, and these systematic errors have to be taken into account if the accuracy potential of the CMM is to be more fully realised. Two approaches can be adopted to nullify the effect of these systematic errors. The first – *error mapping* – involves performing a set of experiments to characterise as completely as possible the error behaviour of the instrument and then use error correction software to produce more accurate co-ordinate estimates. The disadvantages of this approach are, firstly, the set of experiments is expensive to perform and, secondly and more importantly, the error behaviour of the CMM is likely to drift so that, for example, an error correction valid on Monday will only be partially valid on Friday and may be of limited value a month later. The second approach – *self-calibration* – attempts to use any approximate symmetries, rotational or translational, of the artefact so that systematic errors associated with the measuring system are identified as part of the measurement process [4]. The advantage of this method is that the effect of systematic error behaviour of the instrument is cancelled out and the accuracy of the measurements are limited only by the smaller, random component.

### 3.1 Calibration of reference artefacts in 2-dimensions

As an example, we consider the accurate calibration of 2-dimensional artefacts by a two dimensional CMM. The artefacts define the location of targets nominally aligned in a grid pattern. Let $\mathbf{y}_j$, $j = 1, \ldots, n_Y$, be the locations of the targets in a fixed frame of reference, and let

$$\mathbf{y}_{j,k} = T(\mathbf{y}_j, \mathbf{t}_k)$$

be the location of the $j$th target in the $k$th measuring position. Here, the roto-translation $T$ is specified by three parameters $\mathbf{t}$ defining the translation vector and angle of rotation.

We suppose the systematic error of the two dimensional CMM can be expressed as

$$\mathbf{x}^* = \mathbf{x}^*(\mathbf{x}, \mathbf{b}) = \mathbf{x} + \mathbf{e}(\mathbf{x}, \mathbf{b}),$$

where $\mathbf{x}^*$ are the true point co-ordinates, $\mathbf{x}$ are the indicated point co-ordinates output by the machine and $\mathbf{e}(\mathbf{x}, \mathbf{b})$ is the error correction term depending on $\mathbf{x}$ and error parameters $\mathbf{b}$. For instance, suppose the model describes scale and orthogonality errors so that

$$x^* = x(1 + b_1) + y(1 + b_2) \sin b_3, \quad y^* = y(1 + b_2) \cos b_3.$$

If $\mathbf{x}_i$ is the measurement of the $j$th target with the artefact in the $k$th position then the associated observation equation is

$$\mathbf{x}_i + \mathbf{e}(\mathbf{x}_i, \mathbf{b}) = \mathbf{y}_{j,k} + \boldsymbol{\epsilon}_i. \tag{3.1}$$

Given a set of such measurements $\{\mathbf{x}_i\}_1^{m_X}$ and associated index functions $(j(i), k(i))$ specifying the targets and artefact positions, estimates of the model parameters can be determined by solving a non-linear least squares problem

$$\min_{\{\mathbf{y}_j\},\{\mathbf{t}_k\},\mathbf{b}} \quad \sum_{i=1}^{m_X} \mathbf{f}_i^{\mathrm{T}} \mathbf{f}_i,$$

where $\mathbf{f}_i(\mathbf{y}_{j(i)}, \mathbf{t}_{k(i)}, \mathbf{b}) = \mathbf{x}_i + \mathbf{e}(\mathbf{x}_i, \mathbf{b}) - \mathbf{y}_{j,k}$.

The model involves three sets of the parameters: the target locations $\{\mathbf{y}_j\}$, transformation parameters $\{\mathbf{t}_k\}$ and the error parameters $\mathbf{b}$. Each observation equation depends on only one target and one transformation, so that the Jacobian matrix $J$ of partial derivatives can be ordered to have a block-angular structure [2]

$$J = \begin{bmatrix} K_1 & & & J_1 \\ & K_2 & & J_2 \\ & & \ddots & \vdots \\ & & K_{m_X} & J_{m_X} \end{bmatrix},$$

where $K_j$ corresponds to the parameters $\mathbf{y}_j$ and the border blocks $\{J_j\}$ correspond to the border parameters $\mathbf{a} = \{\{\mathbf{t}_k\}, \mathbf{b}\}$. The frame of reference for the targets $\{\mathbf{y}_j\}$ can be specified by applying three appropriate linear equality constraints on the transformation parameters $\{\mathbf{t}_k\}$.

While scale and orthogonality errors are often major contributors to the systematic error behaviour of a CMM, there is no guarantee nor does experience show that they explain the full extent of the behaviour. For this reason, more comprehensive models have been developed [3, 9]. However, they all depend on the approximation of actual behaviour by empirical functions such as polynomials and the adequacy of the approximation is often difficult and expensive to evaluate. However, if we always rotate and translate the artefact according to the symmetries of the reference artefact so that the targets are always located (nominally) at a subset of a fixed grid of points in the CMM's working volume, then measurements are made at a finite number of machine locations. To the $l$th location we associate a machine error $\mathbf{e}_l$. If the $i$th measurement is made at the $l$th location then the observation equation corresponding to (3.1) is

$$\mathbf{x}_i + \mathbf{e}_l = \mathbf{y}_{j,k} + \boldsymbol{\epsilon}_i.$$

The advantage of this error model is that it entails no significant approximation: the

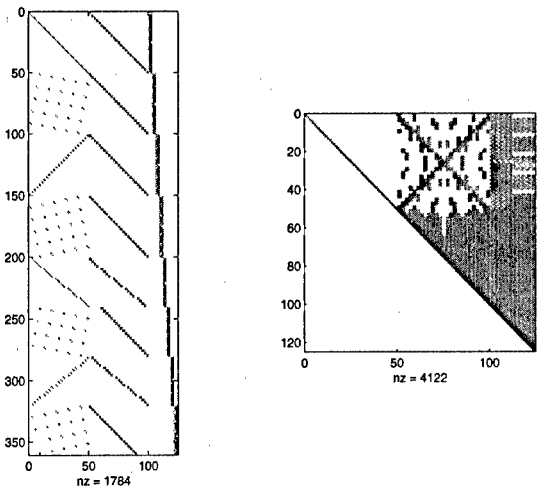


FIG. 1. Sparsity structure of the transpose of the Jacobian matrix associated with the measurement of a $5 \times 5$ hole plate in eight positions.

systematic errors are modelled exactly. An apparent disadvantage is that there are likely to be as many error parameters as target parameters giving rise to a sparsity structure in the Jacobian matrix for which direct, structure-exploiting methods provide relatively minor efficiency gains. Figure 1 shows on the left the sparsity structure of the Jacobian matrix $J$ associated with the measurement of a $5 \times 5$ hole plate in eight positions, the first four corresponding to rotations by 0, 90, 180 and 270 degrees, the second four incorporating a translation as well as a rotation. In each position the location of the targets $\mathbf{y}_j$ are measured in order. The nonzero elements of the matrix are represented by a dot. The first (second) 50 columns correspond to the derivatives with respect to the machine error parameters $\mathbf{e}_l$ (target parameters $\mathbf{y}_j$) and the last 24 correspond to the eight sets of transformation parameters $\mathbf{t}_k$. On the right the sparsity structure of the triangular factor of $J$ is illustrated and shows the substantial fill-in that occurs.

In the next two sections, we describe approaches for dealing efficiently with block-angular and more general sparse-block structure.

## 4 Algorithms for block-angular systems

We consider non-linear least squares problems where the optimisation parameters can be partitioned into two sets $\boldsymbol{\eta} = \{\mathbf{y}_j\}_1^{n_Y}$ and $\mathbf{a}$, and such that each observation equation involves $\mathbf{a}$ and at most one set of parameters $\mathbf{y}_j$. Corresponding to (2.1), we have instead an objective function of the form

$$F(\boldsymbol{\eta}, \mathbf{a}) = \mathbf{f}_0^{\mathrm{T}}(\mathbf{a})\mathbf{f}_0(\mathbf{a}) \quad + \quad \sum_j \mathbf{f}_j^{\mathrm{T}}(\mathbf{y}_j, \mathbf{a})\mathbf{f}_j(\mathbf{y}_j, \mathbf{a}).$$

The associated Jacobian matrix $J$ and its triangular factor $R$ can be arranged to have the form

$$J = \begin{bmatrix} K_1 & & & & J_1 \\ & K_2 & & & J_2 \\ & & \ddots & & \vdots \\ & & & K_{n_Y} & J_{n_Y} \\ & & & & J_0 \end{bmatrix}, \quad R = \begin{bmatrix} R_1 & & & & B_1 \\ & R_2 & & & B_2 \\ & & \ddots & & \vdots \\ & & & R_{n_Y} & B_{n_Y} \\ & & & & B_0 \end{bmatrix}.$$

The nonzero blocks of the matrix $R$ can be stored compactly in a vector $\mathbf{r}$, row by row.

Efficient updating strategies for such triangular factors have been incorporated into a non-linear least-squares solver to deal with block-angular problems. It is assumed that the Jacobian matrix is composed of $n_B$ blocks of rows, with the $i$th block depending on at most one set of parameters $\mathbf{y}_j$, $j = j(i)$. The user is required to supply a function and gradient evaluation module that given $\boldsymbol{\eta}$, $\mathbf{a}$ and $1 \le i \le n_B$, returns $j = j(i)$ and

$$\begin{aligned} \mathbf{f}_i(\mathbf{a}),\ J_i, \quad & j = 0, \\ \mathbf{f}_i(\mathbf{y}_j, \mathbf{a}),\ J_i,\ K_i, \quad & j > 0. \end{aligned}$$

For each $i$, the triangular factor and righthand side vector is updated by the $i$th block of rows:

$$\begin{bmatrix} R_{j(i)} & B_{j(i)} \\ K_i & J_i \end{bmatrix} \mapsto \begin{bmatrix} R_{j(i)} & B_{j(i)} \\ \mathbf{0} & \check{J}_i \end{bmatrix}, \quad \begin{bmatrix} R_0 \\ \check{J}_i \end{bmatrix} \mapsto \begin{bmatrix} R_0 \\ \mathbf{0} \end{bmatrix}.$$

Linear equality constraints on the border parameters $\mathbf{a}$ implemented using the orthogonal projection approach can be incorporated by setting $J_i := J_i V_2$ at the appropriate stage.

## 5 Algorithms for sparse-block matrices

Let $m \times n$ matrix $S$ be composed of $n_B$ submatrices $S_k$ of dimension $m_k \times n_k$. We assume that $S_k$ is stored (column-wise or row-wise) as a column vector $\mathbf{s}_k$. The information in $S$ can be encoded in a column vector $\mathbf{s}_I$ and an indexing set $I_S$ such that $I_S(1:5,k) = (i_k, j_k, m_k, n_k, l_k)$ where $(i_k, j_k)$ specifies the location of $S_k(1,1)$ in $S$ and $l_k$ indicates that $\mathbf{s}_k = \mathbf{s}_I(l_k : l_k + m_k n_k - 1)$. Blocks of such matrices can be easily represented by concatenating the s-vectors and index matrices $I_S$ and performing some trivial index modifications. Matrix-vector multiplications of the form $\mathbf{y} := \alpha S\mathbf{x} + \beta\mathbf{y}$ are easily implemented through a sequence of full matrix multiplications: $\mathbf{y} := \beta\mathbf{y}$, followed by

$$\mathbf{y}(i_k : i_k + m_k - 1) := \mathbf{y}(i_k : i_k + m_k - 1) + \alpha S_k \mathbf{x}(j_k : j_k + n_k - 1),$$

$k = 1, \ldots, n_B$. A similar scheme calculates $\mathbf{x} := \alpha S^{\mathrm{T}}\mathbf{y} + \beta\mathbf{x}$. The storage and multiplication scheme can be modified to take into account the type or structure of the submatrices $S_k$.

To implement linear equality constraints, it is required to perform matrix multiplication by a submatrix $V_2$ of the orthogonal factor of the constraint matrix $C$. A simple scheme can be implemented using the LAPACK routines DGEQRF (orthogonal factorisation) and DORMQR (matrix multiplication by an orthogonal matrix stored as a product of Householder matrices) [8].

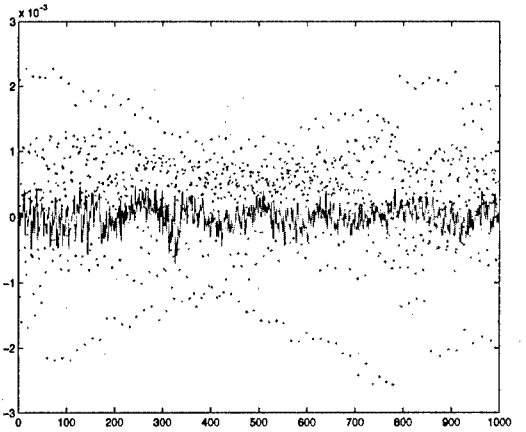


FIG. 2. Residual errors associated with the first 1000 observations for models a) with no error separation (dots) and b) with error separation.

We have implemented a non-linear least squares solver for sparse-block systems. The user is required to supply a module that takes as input the current estimate $\mathbf{a}$ of the optimisation parameters and outputs the function values $\mathbf{f}(\mathbf{a})$ and the Jacobian matrix stored in sparse-block form $\langle \mathbf{s}_I, I_S \rangle$. The solver implements a Gauss-Newton approach using the LSQR solver to find the Gauss-Newton step and caters in a straightforward way for linear equality constraints. The solver has been successfully tested in a number of self-calibration problems. For example, it was used recently in the calibration of a $13 \times 13$ grid of targets on a glass plate by a CMM with an optical probing system. The problem involved approximately 15,000 observation equations in over 800 optimisation parameters and was solved in a few tens of seconds using a standard laboratory PC (450 MHz). The advantage of the error separation model is illustrated in Figure 2 which shows the residual errors associated with the first 1000 observations for models a) with no error separation (dots) and b) with error separation. The fit for the error separation model is much superior. The practical metrological consequence of adopting the enhanced model is that uncertainties associated with the target locations can be reduced by a factor of five. Importantly, because the model is a realistic approximation of the measuring system, we can have confidence in the uncertainty estimates derived from the model.

## 6 Concluding remarks

The move to more accurate measurement systems has led to more comprehensive models of the measuring instrument and its interaction with the physical quantity being measured. These models include parameters that describe properties of the instrument and those of the measurand. The aim of self-calibration experiments is to determine as much as possible about both sets of parameters from a set of measurement experiments. For models with a small to modest set of parameters, a full matrix approach may be acceptable. For larger systems, exploitation of sparsity structure in the defining equations is

highly desirable and often a stark necessity if the computations are to be made in an acceptable time using the computing resources to hand. The exploitation of block-angular structure has been well-known and well-used in some areas of metrology. The supporting numerical technology based on structured orthogonal factorisations is mature, compact and easily implemented using standard numerical linear algebra. However, these techniques could be applied more widely in metrology, making feasible approaches that have to be rejected if full matrix methods only are to be used.

The use of sparse matrix techniques is relatively rare within metrology. We have attempted to show here that in self-calibration problems in dimensional metrology, they allow us to develop improved models that provide vastly superior fits to the data, with corresponding improvements in the evaluated uncertainties in the fitted parameters. The supporting numerical technology is maturing and accessible.

**Acknowledgements:** This work has been supported by the Department of Trade and Industry's National Measurement System Software Support for Metrology Programme and undertaken by a project team at the Centre for Mathematics and Scientific Software, National Physical Laboratory. The author is particularly thankful to Maurice Cox, Peter Harris and Ian Smith for their contributions.

## Bibliography

1. A. Björck. *Numerical Methods for Least Squares Problems.* SIAM, Philadelphia, 1996.
2. M. G. Cox. The least-squares solution of linear equations with block-angular observation matrix. In M. G. Cox and S. Hammarling, editors, *Advances in Reliable Numerical Computation*, pages 227–240. Oxford University Press, 1989.
3. M. G. Cox, A. B. Forbes, P. M. Harris, and G. N. Peggs. Experimental design in determining the parametric errors of CMMs. In V. Chiles and D. Jenkinson, editors, *Laser Metrology and Machine Performance IV*, pages 13–22, Southampton, 1999. WIT Press.
4. A. B. Forbes and I. M. Smith. Self-calibration and error separation techniques in metrology. In P. Ciarlini, M. G. Cox, E. Filipe, F. Pavese, and D. Richter, editors, *Advanced Mathematical and Computational Tools in Metrology V*, pages 149–163, Singapore, 2001. World Scientific.
5. P. E. Gill, W. Murray, and M. H. Wright. *Practical Optimization.* Academic Press, London, 1981.
6. G. H. Golub and C. F. Van Loan. *Matrix Computations.* John Hopkins University Press, Baltimore, third edition, 1996.
7. C. C. Paige and M. A. Saunders. LSQR: and algorithm for sparse linear equations and sparse least squares. *ACM Transactions on Mathematical Software*, 8(1), 1982.
8. SIAM, Philadelphia. *The LAPACK Users' Guide*, third edition, 1999.
9. G. Zhang, R. Ouyang, B. Lu, R. Hocken, R. Veale, and A. Donmez. A displacement method for machine geometry calibration. *Annals of the CIRP*, 37:515–518, 1998.